Inverse problems

- The goal: reconstruct $x$ from $y$
- The common approach until not so long ago:
  Design a specific algorithm for each signal prior and observations model
  - For example, design/apply optimization algorithm
    $$\min_{\tilde{x}} \ell(\tilde{x}, y) + \beta s(\tilde{x})$$
  - Drawback: how to define $s(\tilde{x})$ for sophisticated signals (e.g. natural images)?
Inverse problems

- The goal: reconstruct $x$ from $y$
- The common approach now:
  Design a specific algorithm for each signal prior and observations model
  - Collect/synthesize a training set $\{x_i, y_i\}$ and learn a DNN $f_\theta(y)$ by $\min_\theta \sum_i \|f_\theta(y_i) - x_i\|$
  - Drawback: huge performance drop when the observation model used in training is inexact
Inverse problems

- The goal: reconstruct $x$ from $y$
- In this talk:
  - We show a promising way to balance between the two approaches
    - Fast and simple optimization algorithm that can handle many observation models
    - Enjoying the developments in deep learning to better handle the prior
The observation model

- In many image restoration problems the observations can be formulated by a linear model

\[ y = Ax + e \]

- \( x \in \mathbb{R}^n \) represents the unknown original image
- \( y \in \mathbb{R}^m \) represents the observations
- \( A \in \mathbb{R}^{m \times n} \) is the degradation matrix (known)
- \( e \in \mathbb{R}^m \) is a noise vector (often assumed to be Gaussian)
The observation model

- In many image restoration problems the observations can be formulated by a linear model:

\[ y = Ax + e \]

- Denoising: \( A = I_n \)
- Deblurring: \( A \) is a blurring operator
- Super-resolution: \( A \) composed of blurring and down-sampling
- Other examples: compressed sensing, inpainting, etc.
Typical reconstruction strategy

- Minimize a cost function composed of LS fidelity term and a prior term

\[
\min_{\tilde{x}} \frac{1}{2} \|y - A\tilde{x}\|_2^2 + \beta s(\tilde{x})
\]

- \( s(\tilde{x}) \) is the prior term (regularizer) – required because the measurements do not suffice for a successful recovery (ill-posed inverse problem)

- \( \beta > 0 \) hyper-parameter that controls the level of regularization

- Popular priors: total-variation (convex, explicit), BM3D (non-convex, implicit), etc.
Minimize a cost function composed of LS fidelity term and a prior term

$$\min_{\tilde{x}} \frac{1}{2} \| y - A\tilde{x} \|_2^2 + \beta s(\tilde{x})$$

Optimization methods:
- (Sub-)Gradient descent
  - Requires a lot of iterations (extremely slow)
  - Yields bad local minima for non-convex priors
Typical reconstruction strategy

- Minimize a cost function composed of LS fidelity term and a prior term

\[
\min_{\tilde{x}} \frac{1}{2} \| y - A\tilde{x} \|^2_2 + \beta s(\tilde{x})
\]

- Optimization methods:
  - Variable splitting + ADMM
  - Variable splitting + quadratic penalty method ("HQS")

- More hyper-parameters (for HQS – many more!)
- For non-convex priors: hyper-parameter tuning is burdensome and strongly affects the results
Minimize a cost function composed of LS fidelity term and a prior term

\[
\min_{\tilde{x}} \frac{1}{2} \|y - A\tilde{x}\|_2^2 + \beta s(\tilde{x})
\]

Optimization methods:
- Proximal gradient method ("PGM" / "ISTA")
- Accelerated proximal gradient method ("APGM" / "FISTA")

- Simple
- Can avoid additional hyper-parameter tuning
- However: empirically, for non-convex priors may yield worse results than well-tuned ADMM, HQS (e.g. for deblurring, SR)
“ISTA” on LS + prior cost function

ISTA applied on

$$\min_{\tilde{x}} \frac{1}{2} \|y - A\tilde{x}\|_2^2 + \beta s(\tilde{x})$$

Set step-size $$\mu = \frac{1}{\|A^T A\|}$$ (ensures convergence for convex $$s(\cdot)$$) [Beck and Teboulle, 2009]

Iterate:

$$\tilde{z}_k = \tilde{x}_{k-1} - \mu A^T (A\tilde{x}_{k-1} - y)$$
$$\tilde{x}_k = \text{prox}_{\mu \beta s(\cdot)}(\tilde{z}_k)$$

$$\text{prox}_{s(\cdot)}(\tilde{z}) := \arg\min_{\tilde{z}} \frac{1}{2} \|\tilde{z} - \tilde{x}\|_2^2 + s(\tilde{z})$$ [Moreau, 1965] originally defined for convex $$s(\cdot)$$
“ISTA” on LS + prior cost function

ISTA applied on

\[ \min_{\tilde{x}} \frac{1}{2} \|y - A\tilde{x}\|_2^2 + \beta s(\tilde{x}) \]

Set step-size \( \mu = \frac{1}{\|A^TA\|} \) (ensures convergence for convex \( s(\cdot) \))

Iterate:

\[
\begin{align*}
  \tilde{z}_k &= \tilde{x}_{k-1} - \mu A^T(A\tilde{x}_{k-1} - y) \\
  \tilde{x}_k &= \text{prox}_{\mu \beta s(\cdot)}(\tilde{z}_k)
\end{align*}
\]

\[
\text{prox}_{\mu \beta s(\cdot)}(\tilde{z}) = \arg\min_{\tilde{z}} \frac{1}{2} \|\tilde{z} - \tilde{x}\|_2^2 + \mu \beta s(\tilde{z})
\]

\[
= \arg\min_{\tilde{z}} \frac{1}{2(\sqrt{\mu \beta})^2} \|\tilde{z} - \tilde{x}\|_2^2 + s(\tilde{z}) := \mathcal{D}(\tilde{z}; \sqrt{\mu \beta})
\]

Gaussian denoiser associated with prior \( s(\cdot) \)
“ISTA” on LS + prior cost function

- ISTA applied on \( \min_{\tilde{x}} \frac{1}{2} \|y - A\tilde{x}\|_2^2 + \beta s(\tilde{x}) \)

  Set step-size \( \mu = \frac{1}{\|A^TA\|} \) (ensures convergence for convex \( s(\cdot) \))

  Iterate:
  \[
  \tilde{z}_k = \tilde{x}_{k-1} - \mu A^T (A\tilde{x}_{k-1} - y) \\
  \tilde{x}_k = \mathcal{D}(\tilde{z}_k; \sqrt{\mu \beta})
  \]

- The plug-and-play (P&P) denoisers concept:
  Use **off-the-shelf denoisers** to impose the prior
  (can be done also in ADMM and HQS)

  [Venkatakrishnan et al., Plug-and-play priors, 2013]
Deblurring experiments

- Image deblurring with non-convex P&P priors (BM3D, and pre-trained CNN denoisers):

PSNR vs. iteration number, averaged over 4 scenarios and 8 classical images

Similar behavior (with smaller gap) observed for super-resolution tasks
Deblurring experiments

- Image deblurring with non-convex P&P priors (BM3D, and pre-trained CNN denoisers):

- Our IDBP also uses ISTA... so what makes it faster and better?

[Tirer and Giryes, Iterative Denoising and Back Projections, 2018]
Back-projection (BP) fidelity term

- Typical cost function: LS fidelity + prior
  \[ \min_{\tilde{x}} \frac{1}{2} \| y - A\tilde{x}\|_2^2 + \beta s(\tilde{x}) \]

- Proposed cost function: BP fidelity + prior
  \[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \]

- Assume ill-posed problems: \( m \leq n \) and \( \text{rank}(A) = m \)

- \( A^\dagger := A^T (AA^T)^{-1} \) (“back-projection” from \( A\mathbb{R}^n \) to \( \mathbb{R}^n \))

- LS and BP coincide for denoising and inpainting, but differ for deblurring, super-resolution, Gaussian compressed sensing, etc.
“ISTA” on BP + prior cost function

ISTA applied on
\[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \]

Set step-size \( \mu = \frac{1}{\| A^\dagger A \|} = 1 \) (ensures convergence for convex \( s(\cdot) \))

Iterate:
\[
\begin{align*}
\tilde{z}_k &= \tilde{x}_{k-1} - \mu A^\dagger (A\tilde{x}_{k-1} - y) \\
\tilde{x}_k &= D(\tilde{z}_k; \sqrt{\mu \beta})
\end{align*}
\]
“ISTA” on BP + prior cost function

ISTA applied on
\[
\min_\tilde{x} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x})
\]

Iterate:
\[
\begin{align*}
\tilde{z}_k &= \tilde{x}_{k-1} - A^\dagger (A\tilde{x}_{k-1} - y) \\
\tilde{x}_k &= \mathcal{D}(\tilde{z}_k; \sqrt{\beta})
\end{align*}
\]

We call this method IDBP:
**Iterative Denoising:** \( \tilde{x}_{k-1} = \mathcal{D}(\tilde{z}_{k-1}; \sqrt{\beta}) \)
and **Back-Projections:** \( \tilde{z}_k = \arg\min_{\tilde{z}} \left\| \tilde{z} - \tilde{x}_{k-1} \right\|_2^2 \) s.t. \( A\tilde{z} = y \)
“ISTA” on BP + prior cost function

- ISTA applied on
  \[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \]

Iterate:
\[ \tilde{z}_k = \tilde{x}_{k-1} - A^\dagger (A\tilde{x}_{k-1} - y) \]
\[ \tilde{x}_k = \mathcal{D}(\tilde{z}_k; \sqrt{\beta}) \]

- We call this method IDBP
- The \( A^\dagger \) operation can be performed efficiently by conjugate gradients, FFT (in certain cases), or pre-computation.
BP vs LS fidelity terms – mathematical analysis

- **LS cost:** LS fidelity + prior
  \[
  \frac{1}{2} \| y - A\tilde{x} \|^2_2 + \beta s(\tilde{x}) = \frac{1}{2} \tilde{x}^T A^T A\tilde{x} - y^T A\tilde{x} + c + \beta s(\tilde{x})
  \]

- **BP cost:** BP fidelity + prior
  \[
  \frac{1}{2} \| A^+ y - A^+ A\tilde{x} \|^2_2 + \beta s(\tilde{x}) = \frac{1}{2} \tilde{x}^T A^+ A\tilde{x} - y^T A^+ A^T \tilde{x} + c + \beta s(\tilde{x})
  \]

- **SVD of** \( A = U[\text{diag}\{\lambda_1, \ldots, \lambda_m\}, 0_{n-m}]V^T \)
  - Hessian of LS term: \( A^T A = \sum_{i=1}^{m} \lambda_i^2 v_i v_i^T \)
  - Hessian of BP term: \( A^+ A = \sum_{i=1}^{m} v_i v_i^T \)
  - The “restricted condition number” of BP’s Hessian is better, yet both Hessians are rank deficient

- Implications for MSE and converges rates of algorithms?

---

[Tirer and Giryes, BP Term for Ill-Posed Linear Inverse Problems, 2019]
BP vs LS fidelity terms – mathematical analysis (MSE)

- Assume no noise $y = Ax$
- LS cost: LS fidelity + prior
  \[
  \frac{1}{2} ||Ax - A\tilde{x}||_2^2 + \beta s(\tilde{x}) = \frac{1}{2} (x - \tilde{x})^T A^T A(x - \tilde{x}) + \beta s(\tilde{x})
  \]
  \[
  = \frac{1}{2} \sum_{i=1}^{m} \lambda_i^2 |v_i^T (x - \tilde{x})|^2 + \beta s(\tilde{x})
  \]
- BP cost: BP fidelity + prior
  \[
  \frac{1}{2} ||A^\dagger Ax - A^\dagger A\tilde{x}||_2^2 + \beta s(\tilde{x}) = \frac{1}{2} (x - \tilde{x})^T A^\dagger A(x - \tilde{x}) + \beta s(\tilde{x})
  \]
  \[
  = \frac{1}{2} \sum_{i=1}^{m} |v_i^T (x - \tilde{x})|^2 + \beta s(\tilde{x})
  \]
- Who resembles $||x - \tilde{x}||_2^2 = \sum_{i=1}^{n} |v_i^T (x - \tilde{x})|^2$ more?
BP vs LS fidelity terms – mathematical analysis (MSE)

- Empirical evidence (many experiments and priors: TV, BM3D, DnCNN, DCGAN, RED variants):
  - When the noise level is moderate: BP outperforms LS when the condition number of $AA^T$ is bad. For example: super-resolution, deblurring, some Gaussian CS scenarios
  - If $A$ has very small singular values (SR, deb) at some noise level BP becomes inferior to LS (regularization can help)

Fig. 10: Super-resolution with Gaussian filter and scale factor of 3, using TV prior and 100 iterations of FISTA. PSNR (averaged over 8 test images) vs. $\beta$ (regularization parameter), for (a) $\sigma_e = 0$, and (b) $\sigma_e = \sqrt{2}$.

Fig. 12: Deblurring with uniform $9 \times 9$ blur kernel, using TV prior and 100 iterations of FISTA. PSNR (averaged over 8 test images) vs. $\beta$ (regularization parameter), for (a) $\sigma_e = \sqrt{0.3}$, and (b) $\sigma_e = \sqrt{2}$. 
BP vs LS fidelity terms – mathematical analysis (MSE)

Empirical evidence (many experiments and priors: TV, BM3D, DnCNN, DCGAN, RED variants):

- When the noise level is moderate:
  BP outperforms LS when the condition number of $\mathbf{A} \mathbf{A}^T$ is bad. For example: super-resolution, deblurring, some Gaussian CS scenarios

- If $\mathbf{A}$ has very small singular values (SR, deb) at some noise level BP becomes inferior to LS (regularization can help)

Fig. 17: Super-resolution with Gaussian filter and scale factor of 3, using BM3D prior and 200 iterations of FISTA. PSNR (averaged over 8 test images) vs. $\beta$ (regularization parameter), for (a) $\sigma_e = 0$, and (b) $\sigma_e = \sqrt{2}$.

Fig. 19: Deblurring with uniform $9 \times 9$ blur kernel, using BM3D prior and 200 iterations of FISTA. PSNR (averaged over 8 test images) vs. $\beta$ (regularization parameter), for (a) $\sigma_e = \sqrt{0.3}$, and (b) $\sigma_e = \sqrt{2}$. 
Empirical evidence (many experiments and priors: TV, BM3D, DnCNN, DCGAN, RED variants):

- When the noise level is moderate:
  BP outperforms LS when the condition number of $AA^T$ is bad. For example: super-resolution, deblurring, some Gaussian CS scenarios

- If $A$ has very small singular values (SR, deb) at some noise level BP becomes inferior to LS (regularization can help)
BP vs LS fidelity terms – mathematical analysis (MSE)

- Empirical evidence (many experiments and priors: TV, BM3D, DnCNN, DCGAN, RED variants):
  - When the noise level is moderate: BP outperforms LS when the condition number of $AA^T$ is bad. For example: super-resolution, deblurring, some Gaussian CS scenarios
  - If $A$ has very small singular values (SR, deb) at some noise level BP becomes inferior to LS (regularization can help)

Fig. 1: The (squared) singular values of $A$ applied on a 64×64 image for: (a) SRx3 with 7×7 Gaussian filter ($\frac{\lambda_i^2}{\lambda_{m}^2} = 2.93e3$); (b) blurring with 9×9 uniform filter ($\frac{\lambda_i^2}{\lambda_{m}^2} = 1.46e7$); (c) CS with $m = 0.1n$ Gaussian measurements and Haar basis ($\frac{\lambda_i^2}{\lambda_{m}^2} = 3.63$); (d) CS with $m = 0.5n$ Gaussian measurements and Haar basis ($\frac{\lambda_i^2}{\lambda_{m}^2} = 33.36$).
BP vs LS fidelity terms – mathematical analysis (MSE)

- \[
\min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \quad \text{vs} \quad \min_{\tilde{x}} \frac{1}{2} \left\| y - A\tilde{x} \right\|_2^2 + \beta s(\tilde{x})
\]

- Consider simple Tikhonov regularization \( s(\tilde{x}) = \frac{1}{2} \| D \tilde{x} \|_2^2 \) and SVDs: \( A = U[\text{diag}\{\lambda_i\}, 0]V^T \), \( D^T D = V\text{diag}\{\gamma_i\}V^T > 0 \)

\[
MSE_{BP} = \text{bias}_{BP}^2 + \text{var}_{BP}
\]
\[
\text{bias}_{BP}^2 \triangleq \sum_{i=1}^{m} \left( \frac{\beta \gamma_i^2}{1 + \beta \gamma_i^2} \right)^2 [V^T x]^2_i + \sum_{i=m+1}^{n} [V^T x]^2_i
\]
\[
\text{var}_{BP} \triangleq \sum_{i=1}^{m} \frac{\sigma_e^2}{\lambda_i^2 (1 + \beta \gamma_i^2)^2},
\]
\[
MSE_{LS} = \text{bias}_{LS}^2 + \text{var}_{LS}
\]
\[
\text{bias}_{LS}^2 \triangleq \sum_{i=1}^{m} \left( \frac{\beta \gamma_i^2}{\lambda_i^2 (1 + \beta \gamma_i^2)} \right)^2 [V^T x]^2_i + \sum_{i=m+1}^{n} [V^T x]^2_i
\]
\[
\text{var}_{LS} \triangleq \sum_{i=1}^{m} \frac{\sigma_e^2}{\lambda_i^2 (1 + \beta \gamma_i^2 / \lambda_i^2)^2},
\]

- For same \( \beta \) and \( \lambda_i < 1 \): \( \text{bias}_{BP}^{(i)} < \text{bias}_{LS}^{(i)} \), \( \text{var}_{BP}^{(i)} > \text{var}_{LS}^{(i)} \)

[Tirer and Giryes, BP Term for Ill-Posed Linear Inverse Problems, 2019]
BP vs LS fidelity terms – mathematical analysis (MSE)

- \[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \quad \text{vs} \quad \min_{\tilde{x}} \frac{1}{2} \left\| y - A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \]

- Consider simple Tikhonov regularization \( s(\tilde{x}) = \frac{1}{2} \| D\tilde{x} \|_2^2 \) and SVDs: \( A = U[\text{diag}\{\lambda_i\}, 0]V^T \), \( D^T D = V\text{diag}\{\gamma_i\}V^T > 0 \)

\[
MSE_{BP} = \text{bias}^2_{BP} + \text{var}_{BP} \\
\text{bias}^2_{BP} \triangleq \sum_{i=1}^{m} \left( \frac{\beta \gamma_i^2}{1 + \beta \gamma_i^2} \right)^2 [V^T \underline{x}]_i^2 + \sum_{i=m+1}^{n} [V^T \underline{x}]_i^2 \\
\text{var}_{BP} \triangleq \sum_{i=1}^{m} \frac{\sigma_e^2}{\lambda_i^2(1 + \beta \gamma_i^2)^2} \\
\text{bias}^2_{LS} \triangleq \sum_{i=1}^{m} \left( \frac{\beta \gamma_i^2}{\lambda_i^2 + \beta \gamma_i^2} \right)^2 [V^T \underline{x}]_i^2 + \sum_{i=m+1}^{n} [V^T \underline{x}]_i^2 \\
\text{var}_{LS} \triangleq \sum_{i=1}^{m} \frac{\sigma_e^2}{\lambda_i^2(1 + \beta \gamma_i^2)^2} \\
\text{bias}^{(i)}_{BP} > \text{bias}^{(i)}_{LS}, \quad \text{var}^{(i)}_{BP} < \text{var}^{(i)}_{LS}
\]

[27]

[Tirer and Giryes, BP Term for Ill-Posed Linear Inverse Problems, 2019]
BP vs LS fidelity terms – mathematical analysis (MSE)

- Explanation for the advantage of BP for large $\frac{\lambda_1^2}{\lambda_m^2}$?
- Assume no noise $y = Ax$, then
  $$MSE_{BP} - MSE_{LS} = \sum_{i=1}^{m} bias_{BP}^2(i) - \sum_{i=1}^{m} bias_{LS}^2(i)$$
- Set $\beta_{BP} = \frac{\beta_{LS}}{\lambda_1^2}$
- We get
  $$\sum_{i=1}^{m} bias_{BP}^2(i) = \sum_{i=1}^{m} \left( \frac{\beta_{BP} y_i^2}{1 + \beta_{BP} y_i^2} \right)^2 [V^T x]_i^2$$
  $$= \sum_{i=1}^{m} \left( \frac{\beta_{LS} y_i^2}{\lambda_1^2 + \beta_{LS} y_i^2} \right)^2 [V^T x]_i^2$$
  $$\leq \sum_{i=1}^{m} \left( \frac{\beta_{LS} y_i^2}{\lambda_i^2 + \beta_{LS} y_i^2} \right)^2 [V^T x]_i^2 = \sum_{i=1}^{m} bias_{LS}^2(i)$$
BP vs LS fidelity terms – mathematical analysis (rate)

- \[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \] vs \[ \min_{\tilde{x}} \frac{1}{2} \left\| y - A\tilde{x} \right\|_2^2 + \beta s(\tilde{x}) \]

- Recall: BP requires less (F)ISTA iterations than LS

![Graphs](Fig. 12: Super-resolution with Gaussian filter and scale factor of 3, using BM3D prior. PSNR (for best uniform setting of \( \beta \), averaged over 8 test images) vs. FISTA iteration number, for (a) \( \sigma_c = 0 \), and (b) \( \sigma_c = \sqrt{2} \).)

![Graphs](Fig. 14: Deblurring with uniform 9x9 blur kernel, using BM3D prior. PSNR (for best uniform setting of \( \beta \), averaged over 8 test images) vs. FISTA iteration number, for (a) \( \sigma_c = \sqrt{0.3} \), and (b) \( \sigma_c = \sqrt{2} \).)
BP vs LS fidelity terms – mathematical analysis (rate)

\[ \min_{\tilde{x}} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|^2_2 + \beta s(\tilde{x}) \quad \text{vs} \quad \min_{\tilde{x}} \frac{1}{2} \left\| y - A\tilde{x} \right\|^2_2 + \beta s(\tilde{x}) \]

- Recall: BP requires less (F)ISTA iterations than LS
- We provide theoretical explanation for projected gradient descent (PGD) on the constrained formulations:

\[ \min_{\tilde{x} : s(\tilde{x}) \leq R} \frac{1}{2} \left\| A^\dagger y - A^\dagger A\tilde{x} \right\|^2_2 \quad \text{vs} \quad \min_{\tilde{x} : s(\tilde{x}) \leq R} \frac{1}{2} \left\| y - A\tilde{x} \right\|^2_2 \]

- PGD for BP (with step-size that ensures convergence):

\[ \tilde{x}_{k+1} = \mathcal{P}_{\{\tilde{x} : s(\tilde{x}) \leq R\}} (\tilde{x}_k - A^\dagger (A\tilde{x}_k - y)) \]

- PGD for LS (with step-size that ensures convergence):

\[ \tilde{x}_{k+1} = \mathcal{P}_{\{\tilde{x} : s(\tilde{x}) \leq R\}} (\tilde{x}_k - \frac{1}{\|A^TA\|} A^T (A\tilde{x}_k - y)) \]
BP vs LS fidelity terms – mathematical analysis (rate)

- Toy example: “very restrictive prior”
  \[ s(\tilde{x}) = \begin{cases} 
  0, & \tilde{x}: Q_A \tilde{x} = Q_A x \\
  +\infty, & \text{else} 
  \end{cases} \]

- \[ \mathcal{P}\{\tilde{x}: s(\tilde{x}) \leq R\}(z) = P_A z + Q_A x \]

- For LS: \[ \|\tilde{x}_{k+1} - x_*\|_2 \leq (1 - 1/\text{cond}(AA^T)) \|\tilde{x}_k - x_*\|_2 \]

- For BP: \[ \tilde{x}_{k+1} = A^\dagger y + Q_A x \] (fixed)

- Hints that an advantage of BP may exist even for practical \( s(\tilde{x}) \) which only implicitly impose some restrictions on \( Q_A \tilde{x} \)

\[ Q_A := I_n - P_A \]
\[ P_A := A^\dagger A \]
**Theorem** (informal and simplified, generalizes LS results from [Oymak, Recht, and Soltanolkotabi, 2017]):

Apply PGD on $*\in\{\text{LS, BP}\}$ with $R = s(x)$ (recall, $x$ is the latent signal) and assume no noise. Let $c_s$ be 1 if $s(\cdot)$ is convex and 2 otherwise. Then,

$$
\|\tilde{x}_{k+1} - x\|_2 \leq c_s \cdot P_*(s(\cdot), x) \cdot \|\tilde{x}_k - x\|_2
$$

Let $P_{LS}(s(\cdot), x) := 1 - \frac{1}{\|A^TA\|} \inf_{u \in \mathcal{C}_s(x) \cap S^{n-1}} \frac{1}{2} \|Au\|_2^2$.

Let $P_{BP}(s(\cdot), x) := 1 - \inf_{u \in \mathcal{C}_s(x) \cap S^{n-1}} \frac{1}{2} \|(AA^T)^{-\frac{1}{2}}Au\|_2^2$.

Let $\mathcal{C}_s(x) := \text{cone}\{h \in \mathbb{R}^n : s(x + h) \leq s(x)\}$.
BP vs LS fidelity terms – mathematical analysis (rate)

- **Theorem** (informal and simplified, generalizes LS results from [Oymak, Recht, and Soltanolkotabi, 2017]):
  Apply PGD on \( * \in \{\text{LS, BP}\} \) with \( R = s(x) \) (recall, \( x \) is the latent signal) and assume no noise. Let \( c_s \) be 1 if \( s(\cdot) \) is convex and 2 otherwise. Then,
  \[
  \|\tilde{x}_{k+1} - x\|_2 \leq c_s \cdot P_*(s(\cdot), x) \cdot \|\tilde{x}_k - x\|_2
  \]
  Meaningful if \( c_s P_*(s(\cdot), x) < 1 \) (implies linear convergence) – guaranteed for \( P_{LS} \) in certain scenarios

- **Proposition**: for any full row-rank \( A \) we have
  \[
  P_{BP}(s(\cdot), x) \leq P_{LS}(s(\cdot), x)
  \]
  Implies that whenever PGD converges linearly for LS it also converges for BP (with better or equal rate)
BP vs LS fidelity terms – mathematical analysis (rate)

- The proof technique leads to the conjecture that $P_{BP}(s(\cdot), x) < P_{LS}(s(\cdot), x)$ holds with strict inequality, (at least) for Gaussian compressed sensing (CS) problems.

- CS experiments, $\frac{m}{n} = 0.5$ compression, SNR 20dB

**Projection on $\ell_1$-ball**

**Projection on DCGAN range**
BP vs LS fidelity terms – mathematical analysis (rate)

We also obtained resembling results for general ISTA (subsumes PGD when the prior is convex indicator) if the prox of \( s(\cdot) \) is a contraction on \( Q_A \mathbb{R}^n \)

**Condition B.2.** Given the convex function \( \beta s(\cdot) \) and the full row-rank matrix \( A \), there exists \( 0 < \sigma_{A, \beta s(\cdot)} \leq 1 \) such that

\[
\| \text{prox}_{\beta s(\cdot)}(\tilde{z}_1) - \text{prox}_{\beta s(\cdot)}(\tilde{z}_2) \|_2 \leq \| (P_A + (1 - \sigma_{A, \beta s(\cdot)} Q_A)) (\tilde{z}_1 - \tilde{z}_2) \|_2 \quad \forall \tilde{z}_1, \tilde{z}_2,
\]

where \( Q_A \triangleq I_n - P_A \) and \( P_A \triangleq A^\dagger A \) are the orthogonal projections onto the null space of \( A \) and the row space of \( A \), respectively.

**Theorem B.3.** Consider the penalized optimization problem (2) with convex \( s(\cdot) \) and twice differentiable convex \( \ell(\cdot) \) that satisfies \( \nabla \ell(\cdot) \in \text{range}(A^T) \) for a given full row-rank matrix \( A \). Denote by \( \bar{\sigma}_{max} \) the largest eigenvalue of \( \nabla^2 \ell \) and by \( \bar{\sigma}_{max} \) the smallest non-zero eigenvalue of \( \nabla^2 \ell \). Then, if Condition B.2 holds for \( \mu \beta s(\cdot) \) and \( A \), we have that the sequence \( \{\tilde{x}_t\} \) obtained by (34) with \( \mu = 1/\bar{\sigma}_{max} \) converges to a point \( x_* \) and obeys

\[
\| \tilde{x}_{t+1} - x_* \|_2 \leq \max \left\{ 1 - \frac{\tilde{\sigma}_{min}}{\tilde{\sigma}_{max}}, 1 - \sigma_{A, \beta s(\cdot)} \right\} \| \tilde{x}_t - x_* \|_2,
\]
BP vs LS fidelity terms – mathematical analysis (rate)

We also obtained resembling results for general ISTA (subsumes PGD when the prior is convex indicator) if the prox of \( s(\cdot) \) is a contraction on \( Q_A \mathbb{R}^n \)

Figure 9: Compressed sensing with different \( m/n \) ratios of Gaussian measurements and SNR of 20dB. PSNR (averaged over 4 test images) of PGD with \( \ell_1 \) prior versus iteration number (for \( R = 1.5e5 \)). Note that in these experiments the ratio \( \frac{\sigma_{\min}(AA^T)}{\sigma_{\max}(AA^T)} \) equals 0.0296, 0.0862 and 0.2721 for \( m/n \) ratios of 0.5, 0.3 and 0.1, respectively.
Deblurring experiments

- BSD68 dataset with the following scenarios:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$h(x_1, x_2)$</th>
<th>$\sigma_e^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1/(x_1^2 + x_2^2)$, $x_1, x_2 = -7, \ldots, 7$</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>$1/(x_1^2 + x_2^2)$, $x_1, x_2 = -7, \ldots, 7$</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>$9 \times 9$ uniform</td>
<td>$\approx 0.3$</td>
</tr>
<tr>
<td>4</td>
<td>$[1, 4, 6, 4, 1]^T [1, 4, 6, 4, 1]/256$</td>
<td>49</td>
</tr>
</tbody>
</table>

**Average Deblurring Results (PSNR in dB / SSIM) for Scenarios 1-4 on BSD68 Dataset, and Run-Time (per Image) on Intel i7-7500U CPU @ 2.70 GHz**

<table>
<thead>
<tr>
<th>Method</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Average</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDD-BM3D</td>
<td>30.84 / 0.872</td>
<td>29.02 / 0.820</td>
<td>31.04 / 0.883</td>
<td>28.93 / 0.822</td>
<td>29.96 / 0.849</td>
<td>259s</td>
</tr>
<tr>
<td>P&amp;P-BM3D</td>
<td>30.41 / 0.865</td>
<td>28.53 / 0.806</td>
<td>30.78 / 0.880</td>
<td>28.61 / 0.814</td>
<td>29.58 / 0.841</td>
<td>85s</td>
</tr>
<tr>
<td>IRCNN (≈25 DNNs)</td>
<td>31.17 / 0.877</td>
<td>29.31 / 0.832</td>
<td>30.84 / 0.865</td>
<td>29.16 / 0.830</td>
<td>30.12 / 0.851</td>
<td>34s</td>
</tr>
<tr>
<td>IDBP-BM3D</td>
<td>30.70 / 0.876</td>
<td>28.93 / 0.825</td>
<td>30.80 / 0.883</td>
<td>28.80 / 0.819</td>
<td>29.81 / 0.851</td>
<td>54s</td>
</tr>
<tr>
<td>Auto-tuned IDBP-BM3D</td>
<td>30.75 / 0.872</td>
<td>28.92 / 0.822</td>
<td>30.89 / 0.879</td>
<td>28.74 / 0.821</td>
<td>29.83 / 0.849</td>
<td>152s</td>
</tr>
<tr>
<td>IDBP-CNN (1 DNN per scenario)</td>
<td>31.17 / 0.882</td>
<td>29.19 / 0.830</td>
<td>31.12 / 0.878</td>
<td>29.13 / 0.828</td>
<td>30.15 / 0.855</td>
<td>35s</td>
</tr>
<tr>
<td>Auto-tuned IDBP-CNN (1 DNN per scenario)</td>
<td>31.13 / 0.881</td>
<td>29.18 / 0.828</td>
<td>31.01 / 0.876</td>
<td>29.11 / 0.826</td>
<td>30.11 / 0.853</td>
<td>56s</td>
</tr>
</tbody>
</table>

[Tirer and Giryes, Iterative Denoising and Back Projections, 2018]
Deblurring experiments

Original image

Deblurring of *Barbara* image, Scenario 4
Deblurring experiments

Blurred and noisy image

Deblurring of *Barbara* image, Scenario 4
Deblurring experiments

ADMM-BM3D
Tuned several hyper-params., PSNR=25.72

Deblurring of *Barbara* image, Scenario 4
Deblurring experiments

IDBP-BM3D
Tuned 1 hyper-param., PSNR=26.94

Deblurring of *Barbara* image, Scenario 4
Deblurring experiments

Original image

Deblurring of *cameraman* image, Scenario 3.
Deblurring experiments

Blurred and noisy image

Deblurring of *cameraman* image, Scenario 3.
Deblurring experiments

HQS-CNN (IRCNN)

~25 CNNs, tuned several hyper-params., PSNR=31.07

Deblurring of cameraman image, Scenario 3.
Deblurring experiments

IDBP-CNN

1 CNN, tuned 1 hyper-param., PSNR=31.32

Deblurring of *cameraman* image, Scenario 3.
Super-resolution experiments

“Ideal” observation model:

TABLE I: Super-resolution results (average PSNR in dB) for ideal (noiseless) observation model with bicubic and Gaussian downscaling kernels. Bold black indicates the leading method, and bold blue indicates the leading model-flexible method.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Set5</td>
<td>2</td>
<td>Bicubic</td>
<td>36.66</td>
<td>37.53</td>
<td>38.20</td>
<td>38.27</td>
<td>37.43</td>
<td>37.37</td>
<td>37.41</td>
<td><strong>37.62</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Bicubic</td>
<td>32.75</td>
<td>33.66</td>
<td><strong>34.76</strong></td>
<td>34.74</td>
<td>33.39</td>
<td>33.42</td>
<td>33.44</td>
<td><strong>33.60</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Gaussian</td>
<td>30.42</td>
<td>30.54</td>
<td>30.65</td>
<td>30.74</td>
<td>33.38</td>
<td>31.31</td>
<td>33.48</td>
<td><strong>33.73</strong></td>
</tr>
<tr>
<td>Set14</td>
<td>2</td>
<td>Bicubic</td>
<td>32.42</td>
<td>33.03</td>
<td>34.02</td>
<td><strong>34.12</strong></td>
<td>32.88</td>
<td>33.00</td>
<td>32.95</td>
<td><strong>33.09</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Bicubic</td>
<td>29.28</td>
<td>29.77</td>
<td><strong>30.66</strong></td>
<td>30.65</td>
<td>29.61</td>
<td><strong>29.80</strong></td>
<td>29.65</td>
<td>29.72</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Gaussian</td>
<td>27.71</td>
<td>27.80</td>
<td>27.54</td>
<td>27.80</td>
<td>29.63</td>
<td>28.33</td>
<td>29.68</td>
<td><strong>29.79</strong></td>
</tr>
<tr>
<td>BSD100</td>
<td>2</td>
<td>Bicubic</td>
<td>31.36</td>
<td>31.90</td>
<td>32.37</td>
<td><strong>32.41</strong></td>
<td>31.68</td>
<td>31.65</td>
<td>31.71</td>
<td><strong>31.81</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Bicubic</td>
<td>28.41</td>
<td>28.82</td>
<td><strong>29.32</strong></td>
<td><strong>29.32</strong></td>
<td>28.62</td>
<td>28.67</td>
<td>28.63</td>
<td><strong>28.68</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Gaussian</td>
<td>27.32</td>
<td>27.43</td>
<td>27.46</td>
<td>27.52</td>
<td>28.64</td>
<td>27.76</td>
<td>28.67</td>
<td><strong>28.74</strong></td>
</tr>
</tbody>
</table>

SRCNN, VDSR, EDSR+, RCAN are restricted to their training assumptions (ideal bicubic kernel case). ZSSR is slow (large inference run-time).

[Tirer and Giryes, Super-Resolution via Image-Adapted Denoising CNNs, 2019]
IDBP with image-adaptive CNN denoiser for super-resolution

- Use the IDBP-CNN method with exponentially decreasing noise level, but make the last CNN denoisers - Image Adaptive (IA):
  - Extract patches $\{ p_i \}$ from the input LR image that will serve as ground truth
  - Create their AWGN version with the noise level of the pre-trained denoiser $\{ p_i + \mathcal{N}(0, \sigma I) \}$
  - Fast and simple fine-tuning:
    $$ \min_{\theta} \sum_i \| f_{\theta}(p_i + \mathcal{N}(0, \sigma I)) - p_i \|_1 $$
  - Simple augmentation: random $\{0, 90, 180, 270\}$ rotations and mirror reflections
  - Simple training: fixed 320 iterations of ADAM
  - Optimization time is independent of the image size and the desired SR scale-factor
Super-resolution experiments

“Ideal” observation model:

Fig. 2: Super-resolution results (PSNR averaged on Set5 vs. iteration number) for IDBP-CNN with and without our image-adapted CNN approach: (a) SR x2 with bicubic kernel; (b) SR x3 with Gaussian kernel. A boost in performance is observed once the IDBP scheme starts using image-adapted CNN denoisers.
Super-resolution experiments

- Estimated non-ideal downscaling kernels:

TABLE II: Super-resolution results (average PSNR in dB) for 8 estimated (inexact) non-ideal downscaling kernels and scale factor of 2.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>EDSR+</th>
<th>RCAN</th>
<th>IRCNN</th>
<th>ZSSR</th>
<th>IDBP-CNN</th>
<th>IDBP-CNN-IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set5 (5×8)</td>
<td>29.99</td>
<td>30.01</td>
<td>32.30</td>
<td>33.35</td>
<td>33.33</td>
<td><strong>33.46</strong></td>
</tr>
<tr>
<td>Set14 (14×8)</td>
<td>27.45</td>
<td>27.48</td>
<td>29.24</td>
<td>29.30</td>
<td>29.89</td>
<td><strong>29.96</strong></td>
</tr>
</tbody>
</table>

Fig. 3: (a) Non-ideal downscaling kernels; (b) SR x2 of *monarch*, Set5. From left to right and top to bottom, fragments of: original image, LR image with the estimated kernel, EDSR, RCAN, ZSSR, IDBP-CNN and IDBP-CNN-IA.
Super-resolution experiments

- Real (old) LR images (all methods assume bicubic kernel):

![Images showing different super-resolution methods applied to a basketball player]

- LR
- EDSR+
- ZSSR
- IDBP-CNN
- IDBP-CNN-IA
Super-resolution experiments

- Real (old) LR images (all methods assume bicubic kernel):

<table>
<thead>
<tr>
<th>LR</th>
<th>EDSR+</th>
<th>ZSSR</th>
<th>IDBP-CNN</th>
<th>IDBP-CNN-IA</th>
</tr>
</thead>
</table>

![Real (old) LR images](image-url)
Recovery via generative prior

- We solved inverse problems with CNN denoisers prior
- Another recently popular prior: pre-trained GANs

\[ G_\theta(\cdot) : \mathbb{R}^d \rightarrow \mathcal{X} \subset \mathbb{R}^n \quad (d \ll n) \]

One can constrain the recovery to the range of \( G_\theta(\cdot) \).
Recovery via generative prior

- We solved inverse problems with CNN denoisers prior
- Another recently popular prior: pre-trained GANs
  \[ G_\theta(\cdot) : \mathbb{R}^d \to \mathcal{X} \subset \mathbb{R}^n \ (d \ll n) \]
  One can constrain the recovery to the range of \( G_\theta(\cdot) \)
- CSGM method [Bora et al., 2017]:
  \[ \hat{z} = \arg\min_{z} \| y - AG_\theta(z) \|_2^2 \]
  \[ \hat{x} = G_\theta(\hat{z}) \]
  Minimization w.r.t. \( z \) using GD, ADAM, etc.
- Drawbacks:
  - Non-convex optimization problem
  - **Limited representation capability**
  ("mode collapse")
GANs limited representation capability – examples for PGGAN

- Compressed sensing with $\frac{m}{n} = 0.3$ Fourier measurements

Original image

Naïve IFFT

CSGM
GANs limited representation capability – examples for PGGAN

- Super-resolution x16 with bicubic downscaling kernel

Original image

Bicubic upsampling

CSGM
Image-adaptive GAN (IAGAN)

- IAGAN method (optimize also the generator params.):

\[ \hat{\theta}, \hat{z} = \arg\min_{\theta,z} \| y - AG_\theta(z) \|^2_2 \]

\[ \hat{x} = G_\theta(\hat{z}) \]

Minimize by GD, ADAM, etc.
Use low LR and early stopping to avoid overriding valuable offline semantic information!

- The rationale:
  - Current GAN learning strategies cannot cover every sample of a complex distribution, thus, optimizing only \( z \) is not enough
  - The expressive power of DNNs (given by optimizing the weights \( \theta \) as well) allows to create a single specific sample that agrees with \( y \)
  - Incorporating external and internal learning

Examples for PGGAN

- Compressed sensing with $\frac{m}{n} = 0.3$ Fourier measurements

Original image

Naïve IFFT

DIP

CSGM

IAGAN
Examples for PGGAN

- Super-resolution x16 with bicubic downscaling kernel

Original image

Bicubic upsampling

DIP

CSGM

IAGAN
Compressed sensing experiments

Table 1: Compressed sensing with subsampled Fourier measurements. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for compression ratios 0.3 and 0.5, with noise level of 10/255.

<table>
<thead>
<tr>
<th></th>
<th>CelebA</th>
<th>CelebA-HQ</th>
<th>DIP</th>
<th>CSGM</th>
<th>IAGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS ratio 0.3</td>
<td>19.23 / 0.540</td>
<td>19.65 / 0.625</td>
<td>25.96 / 0.139</td>
<td>24.97 / 0.566</td>
<td>25.50 / 0.092</td>
</tr>
<tr>
<td>CS ratio 0.5</td>
<td>20.53 / 0.495</td>
<td>20.45 / 0.597</td>
<td>27.21 / 0.125</td>
<td>26.29 / 0.535</td>
<td>27.59 / 0.066</td>
</tr>
</tbody>
</table>

![Figure 2: Compressed sensing with Gaussian measurement matrix using BEGAN. Reconstruction MSE (averaged over 100 images from CelebA) vs. the compression ratio \(m/n\).](image)
Super-resolution experiments

Table 2: Super-resolution with bicubic downscaling kernel. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for scale factors 4, 8 and 16, with no noise.

<table>
<thead>
<tr>
<th></th>
<th>CelebA</th>
<th>Bicubic</th>
<th>DIP</th>
<th>CSGM</th>
<th>CSGM-BP</th>
<th>IAGAN</th>
<th>IAGAN-BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR x4</td>
<td>26.50 / 0.165</td>
<td>27.35 / 0.159</td>
<td>20.51 / 0.235</td>
<td>26.44 / 0.165</td>
<td>27.16 / 0.092</td>
<td>27.14 / 0.092</td>
<td></td>
</tr>
<tr>
<td>SR x8</td>
<td>22.39 / 0.212</td>
<td>23.45 / 0.339</td>
<td>20.23 / 0.240</td>
<td>22.71 / 0.212</td>
<td>23.49 / 0.158</td>
<td>23.53 / 0.157</td>
<td></td>
</tr>
<tr>
<td>SR x16</td>
<td>27.43 / 0.437</td>
<td>27.51 / 0.480</td>
<td>22.34 / 0.506</td>
<td>26.20 / 0.437</td>
<td>26.28 / 0.421</td>
<td>25.86 / 0.411</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CelebA-HQ</th>
<th>Bicubic</th>
<th>DIP</th>
<th>CSGM</th>
<th>CSGM-BP</th>
<th>IAGAN</th>
<th>IAGAN-BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR x8</td>
<td>29.94 / 0.398</td>
<td>30.01 / 0.400</td>
<td>22.62 / 0.505</td>
<td>28.54 / 0.398</td>
<td>28.76 / 0.387</td>
<td>28.76 / 0.360</td>
<td></td>
</tr>
<tr>
<td>SR x16</td>
<td>27.43 / 0.437</td>
<td>27.51 / 0.480</td>
<td>22.34 / 0.506</td>
<td>26.20 / 0.437</td>
<td>26.28 / 0.421</td>
<td>25.86 / 0.411</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Super-resolution with bicubic downscaling kernel. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for scale factors 4, 8 and 16, with noise level of 10/255.

<table>
<thead>
<tr>
<th></th>
<th>CelebA</th>
<th>Bicubic</th>
<th>DIP</th>
<th>CSGM</th>
<th>IAGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR x4</td>
<td>24.72 / 0.432</td>
<td>24.19 / 0.280</td>
<td>20.57 / 0.238</td>
<td>25.54 / 0.133</td>
<td></td>
</tr>
<tr>
<td>SR x8</td>
<td>21.65 / 0.660</td>
<td>21.22 / 0.513</td>
<td>20.22 / 0.243</td>
<td>21.72 / 0.243</td>
<td></td>
</tr>
<tr>
<td>SR x16</td>
<td>26.31 / 0.801</td>
<td>27.61 / 0.430</td>
<td>21.60 / 0.519</td>
<td>26.30 / 0.421</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CelebA-HQ</th>
<th>Bicubic</th>
<th>DIP</th>
<th>CSGM</th>
<th>IAGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR x8</td>
<td>26.31 / 0.801</td>
<td>27.61 / 0.430</td>
<td>21.60 / 0.519</td>
<td>26.30 / 0.421</td>
<td></td>
</tr>
<tr>
<td>SR x16</td>
<td>25.02 / 0.781</td>
<td>24.20 / 0.669</td>
<td>21.31 / 0.516</td>
<td>24.73 / 0.455</td>
<td></td>
</tr>
</tbody>
</table>
Take home message

- You can use deep learning for handling only the prior in imaging inverse problems
  - No restrictions on the observation model due to the training phase: the same pre-trained DNNs (denoisers, GANs) are used for different tasks without re-training
  - Exploiting the “good things” in external data: better and faster than methods like DIP that train DNNs from scratch at test time

- BP fidelity term - an alternative to LS
  - Optim. requires less hyper-params and less iterations
  - Mathematical explanation for faster convergence and cases with improved results (see our papers)

- Image-adaptive approach: a simple method to incorporate external and internal learning
Teaser: Robustifying Off-the-Shelf Deep Super-Resolvers

- Inspired by “generalized sampling theory” we apply a correction filter on LR images to mimic bicubic downscaling model (assumed by many DNNs)

Example for SRx4 with Gaussian kernel

Bicubic  proSR  RCAN  DBPN

g.t.
I thank my co-authors

Raja Giryes

Shady Abu Hussein
Thank you

Many experiments and mathematical analysis can be found in:


